



DiscoverySicily Bot



Salvatore Calderaro

Giuseppe Scordato

Indice

1	Descrizione	2
2	Dataset utilizzati	3
2.1	Licenze	3
3	Definizione ontologia	5
4	Pipeline di elaborazione	7
4.1	Unione dei dataset	8
4.2	Interlinking	9
4.3	Creazione del file a cinque stelle	10
4.4	Creazione del bot DiscoverySicily Bot	10

1 Descrizione

DiscoverySicily Bot nasce per permettere ai siciliani e non di conoscere le peculiarità di una città della regione Sicilia. L'utente potrà interagire con il bot in due maniere: inviando la propria posizione oppure inviando un messaggio contenente il nome della città di cui intende conoscere più informazioni. Le informazioni che verranno restituire dal bot riguarderanno le seguenti categorie: musei, castelli, beni immateriali, luoghi del gusto, fortezze e torri.

2 Dataset utilizzati

I dataset utilizzati per la realizzazione del bot sono i seguenti:

- dataset dei comuni italiani: <https://www.istat.it/it/archivio/6789> con licenza *CC-BY-3.0*.
- dataset dei musei siciliani: <https://dati.regione.sicilia.it/dataset/musei-gallerie-e-siti-archeologici> con licenza *CC-BY-4.0*.
- registro beni immateriali: <https://dati.regione.sicilia.it/dataset/registro-eredita-immateriali/resource/de6033fd-783d-424e-aca6-6ef6573693bf> con licenza *CC-BY-4.0*.
- castelli: <https://dati.regione.sicilia.it/dataset/castelli-torri-costiere-forti-e-fortezze/resource/87447c1b-5606-4cce-94e7-f6c897cb53ca> con licenza *CC-BY-SA-4.0*.
- luoghi siciliani del gusto: <https://dati.regione.sicilia.it/dataset/luoghi-siciliani-del-gusto/resource/7f7d39fb-8851-4475-9174-a10069d0e977> con licenza *CC-BY-4.0*.
- torri costiere: <https://dati.regione.sicilia.it/dataset/castelli-torri-costiere-forti-e-fortezze/resource/d0959212-45e3-4cc1-98c7-9e892e8f5ab8> con licenza *CC-BY-SA-4.0*.
- forti e fortezze: <https://dati.regione.sicilia.it/dataset/castelli-torri-costiere-forti-e-fortezze/resource/0efd60e4-c309-407e-a600-ca76fc92eb28> con licenza *CC-BY-SA-4.0*.

Per la geolocalizzazione delle città e dei vari monumenti è stata utilizzata la banca dati di *OpenStreetMap* www.openstreetmap.org con licenza *ODBL*.

2.1 Licenze

I dataset sopra menzionati sono distribuiti con le seguenti licenze:

- *CC-BY-3.0*: licenza di tipo creative commons che permette all’usufruttore di distribuire, modificare e sviluppare anche commercialmente l’opera, riconoscendo sempre l’autore originale;
- *CC-BY-4.0*: licenza di tipo Creative Commons che permette all’usufruttore di distribuire, modificare e sviluppare anche commercialmente l’opera, riconoscendo sempre l’autore originale, presupponendo altresì l’attribuzione automatica di tale licenza nel caso di applicazione del principio “Open Data by default”.

- *CC-BY-SA-4.0*: permette di distribuire, modificare, creare opere derivate dall'originale, anche a scopi commerciali, a condizione che venga riconosciuta una menzione di paternità adeguata, fornito un link alla licenza e indicato se sono state effettuate delle modifiche; e che alla nuova opera venga attribuita la stessa licenza dell'originale (quindi a ogni opera derivata verrà consentito l'uso commerciale)
- *ODBL*: licenza che permette di condividere (copiare, distribuire e utilizzare il database), creare (realizzare opere dal database) e adattare (modificare, trasformare e sviluppare il database) ed obbliga ad attribuire, condividere allo stesso modo e mantenere aperto.

In base alle licenze dei dataset utilizzati il bot ed il relativo dataset verranno rilasciati con licenza *CC-BY-SA -4.0*.

3 Definizione ontologia

Per rappresentare tutte le informazioni inerenti a un comune, abbiamo definito una nostra ontologia *DISCOVERY SICILY ONTOLOGY* (dso). Tale ontologia ha lo scopo di modellare le informazioni peculiari che caratterizzano una città. Inoltre, essendo che le informazioni inerenti quest'ultima (musei, beni immateriali, castelli etc) sono molteplici e di vario tipo, è stato necessario l'utilizzo di tre diverse ontologie già esistenti, in particolare:

- *DBPEDIA* (dbo) per quanto concerne le informazioni di base inerenti la città (nome, provincia, geolocalizzazione);
- *CULTURAL ON* (cis): per quanto riguarda i musei;
- *ONTOLOGIA DELLA REGIONE SICILIANA* (rso): per i musei che non sono rappresentabili mediante *CULTURAL ON*, i castelli, i beni immateriali, i luoghi del gusto, e le torri.

Ogni risorsa rappresentabile dalla nostra ontologia ha oppure potrebbe avere le seguenti proprietà:

- *comune* contenente le informazioni sul nome della città, la provincia e la geolocalizzazione:
 - *dbo:name*: nome della città;
 - *dbo:metropolitanCity* o *dbo:Provincie*: nome della città metropolitana o della provincia;
 - *dbo:lat*: latitudine;
 - *dbo:long*: longitudine.
- *museo*: contenente i musei della città le cui informazioni sono presenti su *CULTURAL ON*:
 - *cis:institutionalCISName*: nome istituzionale del museo;
 - *dc:type*: tipo di museo;
 - *l0:description*: descrizione;
 - *geo:lat*: latitudine;
 - *geo:long*: longitudine.
- *museorso*: contenente i musei non presenti su *CULTURAL ON*, ma nell'Ontologia della Regione Siciliana:

- *rdfs:label*: nome del museo;
 - *rso:hasCulturalPropertyType*: tipo;
 - *rso:hasFeatureLocation*: indirizzo.
- *castello*: contenente le seguenti informazioni inerenti un castello:
 - *rdfs:label*: nome del castello;
 - *rdfs:comment*: breve commento sul castello;
 - *rso:hasDating*: datazione storica del castello;
 - *rso:lat*: latitudine;
 - *rso:lon*: longitudine.
 - *rso:webPageLink*: sito web del castello.
 - *beneimmateriale*: contenente le seguenti informazioni sui beni immateriali:
 - *rdfs:label*: nome della festa;
 - *rso:hasCulturalPropertyType* : tipo di festa;
 - *rso:webPageLink*: link alla quale è possibile reperire più informazioni.
 - *luogo del gusto*: contenente le informazioni sulle particolarità culinarie di un determinato luogo:
 - *rdfs:label*: nome della particolarità culinaria;
 - *rdfs:comment*: breve descrizione della particolarità culinaria;
 - *torre*: contenente le seguenti informazioni su torri, forti, fortezze e torri costiere:
 - *rdfs:label*: nome della torre;
 - *rso:lat*: latitudine;
 - *rso:lon*: longitudine.

t

4 Pipeline di elaborazione

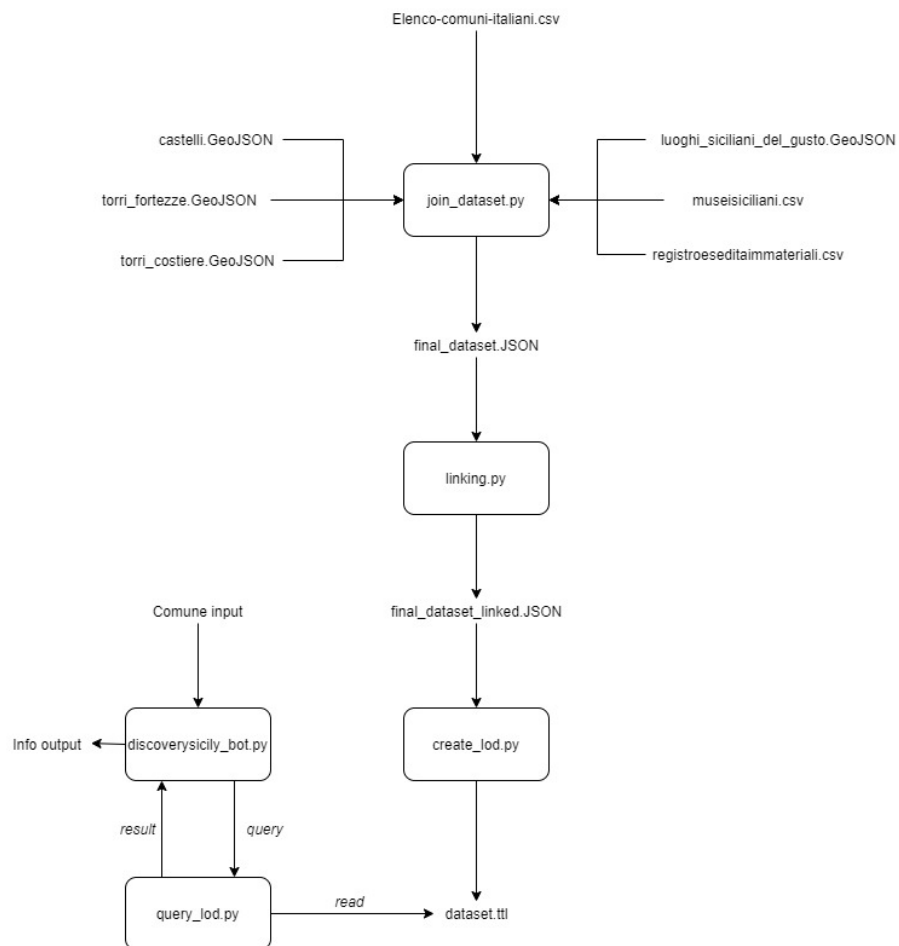


Figura 1: Pipeline di elaborazione

4.1 Unione dei dataset

In una prima fase è stato effettuato il join dei dataset menzionati precedentemente in modo da ottenere per ogni comune le informazioni di nostro interesse. In particolar modo prima sono stati estratti tutti comuni siciliani e le relative informazioni, successivamente mediante delle apposite funzioni abbiamo associato ad ogni comune - ove presenti - musei, castelli, beni immateriali, luoghi del gusto e torri. Per effettuare alcune di queste associazioni è stato necessario effettuare un partial matching tra il nome del comune prelevato dal dataset dei comuni e quello del dataset che si stava analizzando. Un esempio è il seguente:

```
def return_luoghi_gusto(comune, luoghi_gusto):
    lista_luoghi_gusto=[]
    for luogo_gusto in luoghi_gusto:
        aus={}
        try:
            score=token_set_ratio(luogo_gusto[1].lower(),comune.lower())
            if score > 80:
                aus['Coordinate']=luogo_gusto[0]
                aus['Elemento']=luogo_gusto[3]
                aus['Note']=luogo_gusto[4]
                lista_luoghi_gusto.append(aus)
        except:
            pass
    return lista_luoghi_gusto
```

Figura 2: Esempio di partial matching

Ogni comune è stato rappresentato mediante un dizionario avente le seguenti chiavi:

```
comune={

    "Nome Comune":

    "Provincia":

    "Coordinate":

    "Musei":

    "Beni Immateriali":

    "Castelli":

    "Luoghi del gusto":

    "Torri costiere":

    "Torri fortezze":

}
```

Figura 3: Esempio dizionario comune

Alcuni campi sopra citati hanno come valori liste di dizionari, perché ad esempio in una città possono essere presenti più musei etc. L'output di questa prima fase è una lista di dizionari che è stata salvata all'interno di un file JSON (*final_dataset.JSON*).

4.2 Interlinking

Per rendere il nostro dataset linked abbiamo collegato le informazioni inerenti ai comuni ed ai musei rispettivamente alla base di conoscenza di dbpedia e di *dati.beniculturali.it*. La query utilizzata per ottenere i comuni e le relative informazioni da dbpedia è la seguente:

```
res_comuni_dbo = query_sparql("https://dbpedia.org/sparql",
    """
    SELECT DISTINCT ?citta ?label WHERE {
        ?citta a dbo:Location ;
        dbo:country dbr:Italy .
        ?citta rdfs:label ?label.
        FILTER( LANG( ?label) = 'it')
    }
    """)
```

Figura 4: Query comuni dbpedia

mentre quella per ottenere i musei da *dati.beniculturali.it* è la seguente:

```
res_musei_cis = query_sparql("https://dati.beniculturali.it/sparql",
    """
    select distinct ?risorsa ?categoria ?nome ?descrizione ?nome_comune ?indirizzo ?latitudine ?longitudine where {
        ?risorsa a cis:CulturalInstituteOrSite .
        {
            ?risorsa rdf:type cis:ArchaeologicalPark .
        }
        UNION
        {
            ?risorsa rdf:type cis:Museum .
        }
        ?risorsa rdfs:label ?nome .
        ?risorsa geo:lat ?latitudine .
        ?risorsa geo:long ?longitudine .
        ?risorsa dc:type ?categoria .
        ?risorsa l0:description ?descrizione .
        ?risorsa cis:hasSite ?sito .
        ?sito cis:siteAddress ?address .
        ?address clvapit:fullAddress ?indirizzo .
        ?address clvapit:hasRegion <http://dati.beniculturali.it/mibact/luoghi/resource/Region/Sicilia> .
        ?address clvapit:hasCity ?comune .
        ?comune rdfs:label ?nome_comune .
        filter (lang(?descrizione)="it") .
    }
    """
```

Figura 5: Query musei dati.beniculturali.it

In entrambi casi è stato necessario effettuare un partial matching fra i nomi dei comuni presenti del nostro dataset e quelli risultanti dalla query. La medesima cosa è stata fatta per i musei per quanto riguarda la denominazione. Inoltre analizzando approfonditamente i risultati della query sui musei è

emerso che alcuni musei risultati dalla query non erano presenti nel nostro dataset. Alla luce di ciò abbiamo deciso di agire come segue:

- se il museo è presente sia nel nostro dataset e sia nei risultati della query, si prelevano le informazioni sul museo ottenute dalla query in quanto risultano essere più complete e viene inoltre aggiunto un campo `cis_museo` che ci consentirà successivamente di capire se i dati di quel determinato museo provengono dall'ontologia dei beni culturali o meno.
- se come output della query vi è un museo non presente nel nostro dataset, si prelevano le informazioni di quest'ultimo e si aggiungono al dataset.
- per i musei presenti nel nostro dataset, ma che non vengono restituiti dalla query non viene effettuata alcuna operazione in quanto quest'ultimi verranno gestiti mediante l'ontologia della regione siciliana.

Per le altre informazioni non è stato effettuato alcun linking in quanto quest'ultime sono modellabili mediante l'ontologia della regione siciliana. Ciò che si ottiene in output dopo questa fase è il file *final_dataset_linked.JSON*.

4.3 Creazione del file a cinque stelle

Per la creazione del file a cinque stelle è stata importata ed analizzata la lista di dizionari presenti nel file *final_dataset_linked.JSON* e mediante la libreria *rdflib* è stato generato il file *dataset.ttl*. Inoltre in questa fase è stato necessario creare la uri della nostra risorsa: <http://www.discoverysicily.it/resource/Comune/>.

4.4 Creazione del bot DiscoverySicily Bot

Per la realizzazione del bot è stata utilizzata la libreria Python *telepot* che tramite appositi metodi permette di realizzare la comunicazione con i server di Telegram ed interagire con gli utenti connessi. I file utilizzati per mostrare tutte le informazioni sui comuni sono: *final_dataset_linked.JSON* e *dataset.ttl*. Un utente può interagire con il bot in due maniere: inviando la propria posizione oppure mandando un messaggio contenente il nome del comune. Tramite un'apposita funzione viene controllato se i dati inseriti fanno riferimento ad un comune siciliano o meno. In caso affermativo verrà restituito un bottone con il nome della città oppure le città limitrofe viceversa un messaggio di errore.



Figura 6: Ricerca del comune mediante posizione



Figura 7: Ricerca del comune mediante posizione

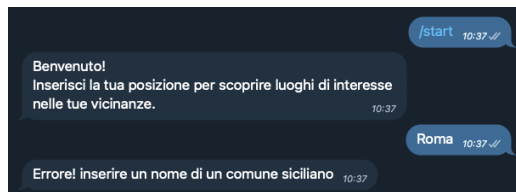


Figura 8: Esempio di inserimento di un comune non siciliano

Una volta cliccato sul bottone della città di cui si vogliono visualizzare le informazioni, mediante delle apposite query SPARQL viene verificato se per quella città sono presenti musei, castelli, beni immateriali, luoghi del gusto e torri. Per ciascuno di queste risorse - se presenti - verrà creato un apposito bottone. Se per la città non viene trovata alcuna risorsa verrà restituito un apposito messaggio. Le query utilizzate sono le seguenti:

```
select distinct ?label ?type ?lat ?long ?description where{
  "" + comune + "" a dbo:Location ;
  cis:Museum ?link_museo .
  ?link_museo cis:institutionalCISName ?label .
  ?link_museo dc:type ?type .
  ?link_museo geo:lat ?lat .
  ?link_museo geo:long ?long .
  ?link_museo l0:description ?description .
}
```

Figura 9: Query per la ricerca dei musei

```
select distinct ?label ?type ?address where{
  "" + comune + "" a dbo:Location ;
  rso:CulturalProperty ?link_museo .
  ?link_museo rdfs:label ?label .
  ?link_museo rso:hasCulturalPropertyType ?type .
  ?link_museo rso:hasFeatureLocation ?address .
}
```

Figura 10: Query per la ricerca dei musei

```

select distinct ?label ?comment ?date ?lat ?long ?webpage where{
  "" + comune + "" a dbo:Location ;
  rso:Castle ?link_castelli .
  ?link_castelli rdfs:label ?label .
  ?link_castelli rdfs:comment ?comment .
  ?link_castelli rso:hasDating ?date .
  ?link_castelli rso:lat ?lat .
  ?link_castelli rso:lon ?long .
  ?link_castelli rso:webPageLink ?webpage .
}

```

Figura 11: Query per la ricerca dei castelli

```

select distinct ?label ?type ?webpage where{
  "" + comune + "" a dbo:Location ;
  rso:IntangibleCulturalProperty ?link_beni .
  ?link_beni rdfs:label ?label .
  ?link_beni rso:hasCulturalPropertyType ?type .
  ?link_beni rso:webPageLink ?webpage .
}

```

Figura 12: Query per la ricerca dei beni immateriali

```

select distinct ?label ?comment where{
  "" + comune + "" a dbo:Location ;
  rso:DemoEthnoAnthropologicalHeritage ?link_gusto .
  ?link_gusto rdfs:label ?label .
  ?link_gusto rdfs:comment ?comment .
}

```

Figura 13: Query per la ricerca dei luoghi del gusto

```

select distinct ?label ?lat ?long where{
  "" + comune + "" a dbo:Location ;
  rso:PointOfInterest ?link_torri .
  ?link_torri rso:lat ?lat .
  ?link_torri rso:lon ?long .
  ?link_torri rdfs:label ?label .
}

```

Figura 14: Query per la ricerca delle torri

Di seguito alcuni esempi di funzionamento:

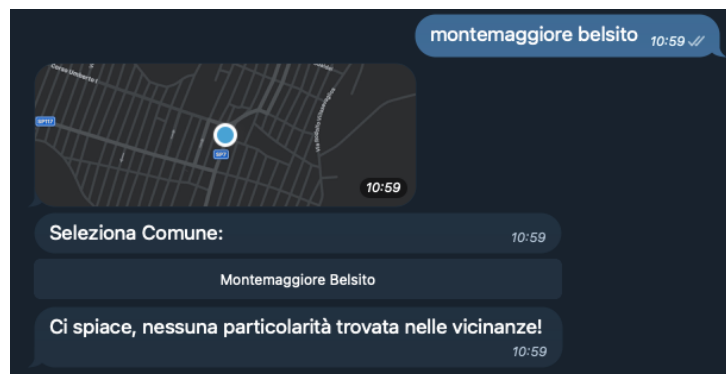


Figura 15: Esempio di comune per cui non è presente alcuna informazione.

Cliccando su uno dei bottoni che vengono restituiti, mediante delle apposite funzioni verranno inviati dei messaggi con le informazioni complete inerenti la risorsa su cui si vuole approfondire:

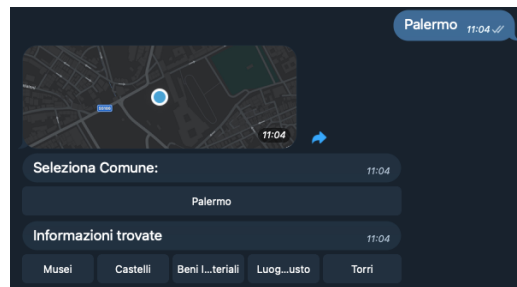


Figura 16: Esempio di comune per cui sono presenti tutte le risorse

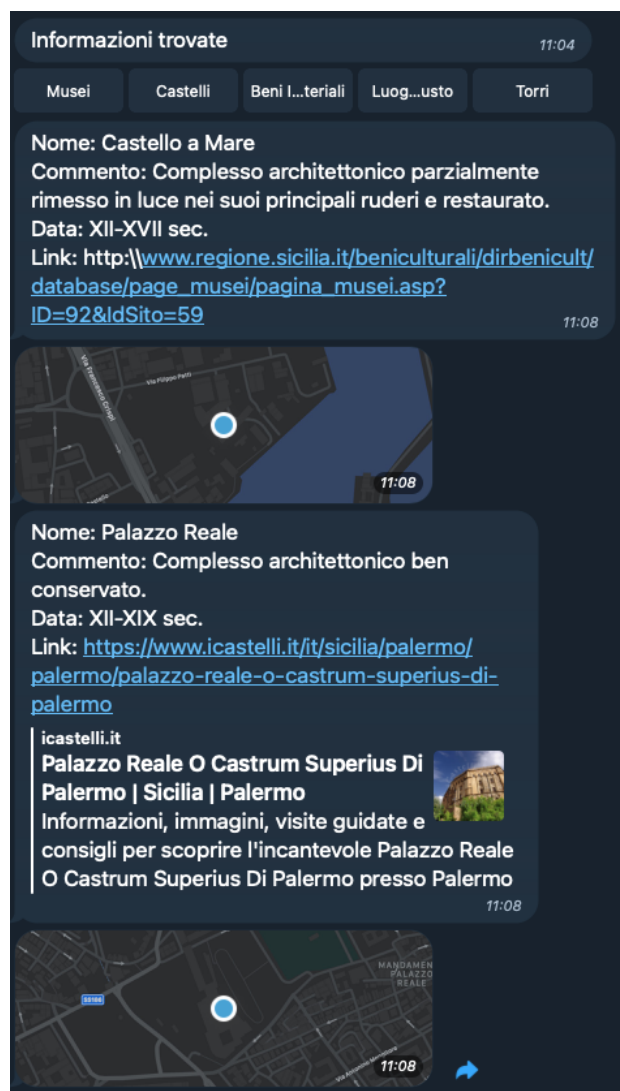


Figura 17: Esempio messaggio con dettagli sui castelli